Available online at www.sciencedirect.com
SCIENCE DIRECT®
Genomics 83 (2004) 148–152
GENOMICS
www.elsevier.com/locate/ygeno

Physical and linkage mapping of mammary-derived expressed sequence tags in cattle

E.E. Connor,[a,*] T.S. Sonstegard,[a] J.W. Keele,[b] G.L. Bennett,[b] J.L. Williams,[c] R. Papworth,[c] C.P. Van Tassell,[a] and M.S. Ashwell[a]

[a] *U.S. Beltsville Agricultural Research Center, ARS, U.S. Department of Agriculture, 10300 Baltimore Avenue, Beltsville, MD 20705, USA*
[b] *U.S. Meat Animal Research Center, ARS, U.S. Department of Agriculture, P.O. Box 166, Clay Center, NE 68933-0166, USA*
[c] *Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, Scotland, United Kingdom*

Received 2 June 2003; accepted 5 July 2003

Abstract

This study describes the physical and linkage mapping of 42 gene-associated markers developed from mammary gland-derived expressed sequence tags to the cattle genome. Of the markers, 25 were placed on the USDA reference linkage map and 37 were positioned on the Roslin 3000-rad radiation hybrid (RH) map, with 20 assignments shared between the maps. Although no novel regions of conserved synteny between the cattle and the human genomes were identified, the coverage was extended for bovine chromosomes 3, 7, 15, and 29 compared with previously published comparative maps between human and bovine genomes. Overall, these data improve the resolution of the human–bovine comparative maps and will assist future efforts to integrate bovine RH and linkage map data.

Keywords: RH mapping; Linkage mapping; SNP; Cattle; EST

Selection of positional candidate genes controlling economically important traits in cattle requires a detailed map of the bovine genome. A recent census of publicly available map information shows that only 1599 genes have been positioned on the various maps of the bovine genome (http://locus.jouy.inra.fr/cgi-bin/bovmap/intro.pl). Although numerous mapping studies have identified genomic regions containing quantitative trait loci (QTL) of interest in dairy cattle, the low resolution of current bovine gene maps is insufficient to select efficiently genes meeting positional and functional criteria and classify them as candidates contributing to these traits. To apply the human sequence or mapping information to candidate gene identification in cattle will also require improving the current comparative map to identify the region of conserved synteny corresponding to the location of QTL [1]. Comparative maps of expressed sequence tags (EST) have been developed for a number of species, including rat [2], mouse [3], pig [4,5], and cattle [6], and serve as a resource for candidate gene identification.

Recent efforts aimed at constructing more comprehensive comparative maps between the cattle and the human genomes have greatly improved the overall comparative coverage and resolution of gene loci order [6–8]. However, these resources need further improvement to utilize fully genome sequence information from other species to identify functional and positional candidate genes underlying QTL. Many of the genes positioned on the various bovine maps lack supporting annotation information that can be correlated to gene expression and function in cattle. For these reasons, we screened EST information from the Institute for Genomic Research (TIGR) *Bos taurus* Gene Index Release 6.0 (http://www.tigr.org/tdb/tgi/btgi/) to identify tentative consensus (TC) sequences that were amenable to integration of the physical and linkage maps by allowing both radiation hybrid (RH) mapping and linkage analysis, the latter via detection of single nucleotide polymorphisms (SNP). Criteria for selection of candidate sequences for mapping included: (1) annotation for human map position and gene function, (2) alignment of the bovine sequence to the human genome such that the alignment was disrupted by

* Corresponding author. Bovine Functional Genomics Laboratory, USDA-ARS, Building 200, Room 4, BARC-East, Beltsville, MD 20705, USA. Fax: +1-301-504-8414.
E-mail address: econnor@anri.barc.usda.gov (E.E. Connor).

0888-7543/$ - see front matter.
doi:10.1016/S0888-7543(03)00218-0

putative human intron sequence, and (3) that the bovine TC sequence comprised >75% mammary-derived EST. The first two criteria were similar to the selection strategy used by Stone and colleagues [6] in positioning genes on the reference linkage map using EST information. The latter criterion would provide map information for those genes that appeared relatively specific to, and important for, mammary gland development and function based on the current EST information available for cattle. Similar approaches have been applied for mapping tissue-specific EST in humans [9,10] and swine [11–13]. In the present study, 60 mammary-expressed genes were selected for mapping, and markers representing 42 of these genes were positioned on the physical and linkage maps.

Table 1
Summary of physical and linkage mapping of 46 bovine gene-associated STS

BTA	Locus	STS Accession Nos. (SNP)[a]	Linkage map	RH map	HSA ($\times 10^6$ bp)[b]
			marker (2-point LOD)	marker (2-point LOD)	
1	S100B	BV005797 (285), BV005798	BMS599 (23.18)	INRA250 (9.63)	21 (44.53)
2	ATIC	BV005758, BV005759	ND[c]	ATICR (11.70)	2 (214.36)
2	UBXD2	BV005809 (250), BV005810	BMS1126 (11.14)	BMS1126 (6.25)	2 (134.82)
2	WDR12	BV005814 (567)	BMS778 (13.2)	TGLA226 (7.09)	2 (201.98)
3	INPP5B	BV005777, BV005778 (101)	BM7225 (12.79)	INRA200 (5.30)	1 (37.42)
3	PGM1	BV005794	ND	BM3020 (6.73)	1 (64.02)
3	SHC1	BV005799 (318)	ILSTS096 (5.47)	BMS2904 (13.59)	1 (148.19)
4	GARS	BV005775	ND	BMS2571 (7.84)	7 (30.30)
5	CHD4	BV005769 (291)	BMS1658 (29.61)	CD9 (11.06)	12 (6.55)
5	KRT5	BV005780, BV005781	ND	BMC1009 (17.71)	12 (53.05)
5	LALBA	BV005782	ND	CSSM034 (16.44)	12 (48.93)
5	PACSIN2	BV005792 (418)	ETH2 (9.58)	BMS597 (11.06)	22 (39.94)
5	PZP	BV005796 (278)	BM315 (6.36)	ND	12 (9.04)
5	SUOX	BV005802	ND	ETH10 (4.63)	12 (56.18)
7	ALDH7A1	BV005757 (32)	TGLA303 (10.24)	TGLA303 (9.67)	5 (125.54)
7	C3	BV005764, BV005765	ND	CSSM020 (6.73)	19 (6.79)
7	HSPA9B	BV005776	ND	CSSM057 (5.49)	5 (137.78)
7	SLC22A4		ND	IL4 (8.29)	5 (131.18)
8	B4GALT1	BV005761 (174)	ND	BMS381 (4.65)	9 (33.45)
8	EPB72	BV005773 (968,1097,1249)	BMS2629 (11.87)	BMS2629 (8.31)	9 (115.22)
11	CRAT	BV005770	ND	ETH9 (10.93)	9 (122.98)
11	MAT2A	BV005788 (277)	ILSTS036 (7.97)	BM6445 (10.03)	2 (85.66)
11	PAPOLG	BV005793 (197)	BMS1716 (15.22)	ILSTS100 (18.65)	2 (61.13)
13	ENTPD6	BV005771, BV005772	HUJ616 (15.65)	BMS1580 (17.15)	20 (25.18)
13	LBP	BV005783, BV005784 (375)	BM9248 (8.26)	ABS10 (11.90)	20 (36.72)
15	CASP4	BV005766 (575,841)	MGTG13B (5.27)	ND	11 (107.02)
15	CAT	BV005767 (37), BV005768	POTCHA (9.63)	ILSTS27 (10.76)	11 (36.03)
15	MTMR2	BV012664 (118,133,202,235,419,525)	BMS2533 (25.76)	BR3510 (9.63)	11 (97.12)
18	BAX	BV005763 (11,219,261)	BM2078 (19.07)	ND	19 (50.12)
19	EZHI	BV005774	ND	IOBT965 (15.66)	17 (40.44)
19	LPO	BV005785, BV05786 (91)	HEL10 (16.52)	LPO (13.28)	17 (48.06)
19	TIMP2	BV005806, BV005807 (28)	CIOBT34 (8.13)	ND	17 (76.63)
19	USP22	BV005811 (174)	BMS650 (18.20)	BMS650 (12.59)	17 (21.94)
22	LTF	BV005787 (209,210,219,220,324,332)	DIK115 (9.33)	BMS1932 (21.74)	3 (45.73)
23	TAP1	BV005803 (273)	BM47 (10.83)	LMP7 (13.29)	6 (32.81)
25	AZGP1	BV005760	ND	ARO4 (11.72)	7 (98.10)
26	ACTR1A	BV005756	ND	INRA081 (6.73)	10 (103.14)
28	PLAU	BV005755 (236)	BMS362 (14.75)	BMS362 (19.17)	10 (74.63)
29	ITM1	BV005779	ND	BY1505 (13.94)	11 (127.48)
29	ST14	BV005800 (293)	BL1100 (8.43)	CD5 (11.63)	11 (132.06)
29	THRSP	BV005804 (34,392,462) BV005805 (237,379,436,604,642)	RM179 (6.62)	ND	11 (80.10)
X	MSL3L1	BV005789	ND	IDVGA82 (6.66)	X (10.87)
ND	STIP1	BV005801	ND	ND	11 (66.41)
ND	VCL	BV005812, BV005813	ND	ND	10 (74.77)
ND	MSN	BV005790, BV005791	ND	ND	X (61.15)
ND	TSN	BV005808	ND	ND	2 (119.72)

[a] dbSTS (http://www.ncbi.nlm.nih.gov). Number in parentheses indicates position of SNP in the dbSTS reference sequence used for linkage mapping.
[b] Human Genome Working Draft (http://genome.ucsc.edu/). Number in parentheses indicates approximate number of bases in chromosome from p arm telomere.
[c] ND, not determined.

Results and discussion

Primer pairs were designed to PCR-amplify 60 genes expressed in the mammary gland for physical and linkage mapping. Initially, uniform PCR conditions were tested for all 60 genes. Ultimately, 46 primer pairs amplified targeted genes of interest, but required 19 unique sets of PCR conditions. Modifications included variations in concentration of DNA template, thermal cycling conditions, source of *Taq* DNA polymerase, and use of commercial additives for PCR optimization. Conditions used for PCR are described with each entry in the GenBank database of sequence tagged sites (dbSTS).

Table 1 summarizes the mapping of 46 gene-associated STS on the Roslin 3000-rad RH map and the USDA linkage map and provides the position of the human ortholog within the June 2002 Human Genome Draft Sequence (http://genome.ucsc.edu/). Of the 46 gene-associated STS amplified, 28 primer pairs produced unique products suitable for RH mapping. Nested primers were designed for an additional 9 gene-associated STS to produce bovine-specific amplification products from the RH panel. Two gene-associated STS (*B4GALT1* and *SUOX*) placed on the RH map produced a lod score less than 5.0. However, the assignment of *B4GALT1* to BTA8 is consistent with the findings of Karall-Albrecht and colleagues [14] and positioning of *SUOX* on BTA5 is predicted based on the current human–bovine comparative map. Twenty-five STS exhibited one or more SNP informative in at least one of the four MARC reference sires that were used for linkage mapping. In total, 20 gene-associated STS (*ALDH7A1*, *CAT*, *CHD4*, *ENTPD6*, *EPB72*, *INPP5B*, *LBP*, *LTF*, *MAT2A*, *MTMR2*, *PACSIN2*, *PAPOLG*, *PLAU*, *S100B*, *SHC1*, *ST14*, *TAP1*, *UBXD2*, *USP22*, and *WDR12*) were positioned on both bovine physical and linkage maps, with no discrepancy in chromosomal assignment between the two maps. Placement of *INPP5B* and *SHC1* (BTA3), *HSPA9B* (BTA7), *ITM1* and *MTMR2* (BTA15), and *THRSP* (BTA29) on the bovine maps extended the current confirmed conserved regions of synteny between the human genome and the bovine chromosomes 3, 7, 15, and 29 by approximately 1.2, 0.8, 7.0, and 8.0 Mb, respectively (Fig. 1).

From approximately 37 kb of bovine sequence obtained, 49 SNP were confirmed in the MARC mapping population and used for linkage mapping. An additional 25 putative SNP were identified by initial sequencing, but were not validated further in the MARC reference population. One gene-associated STS (*ENTPD6*) exhibited a dinucleotide (GT) repeat and five STS showed the presence of insertion/deletions. Thus, our approach was successful in identifying SNP markers suitable for mapping in 54% of genes amplified from which an average of 822 bp of sequence was obtained per gene-associated STS.

Four gene-associated STS (*STIP1*, *VCL*, *MSN*, and *TSN*) could not be positioned on either the RH or linkage maps. Primers designed to amplify *STIP1* amplified a 200-bp product that did not span an intron. No SNP were detected in the 200-bp product and a product amplified from hamster genomic DNA was indistinguishable from bovine DNA, precluding the use of the RH panel for physical mapping. Primers designed to amplify *VCL* amplified the targeted region from bovine genomic DNA spanning an intron (1.8 kb) that was distinguishable from hamster (1.6 kb); however, no cell lines in the RH panel appeared to have retained this region of the bovine genome and no SNP were detected in 1126 bp of the amplicon sequenced. Similarly, no SNP were identified for *MSN* (1174 bp evaluated) and PCR amplification was not suitable for RH mapping. Last, PCR amplification of *TSN* produced spurious results from the RH panel and results of sequencing analysis of the amplified fragment suggested a gene duplication. Thus genotypes could not be interpreted for linkage mapping. Table 1 provides the accession numbers of these four gene-associated STS and the position of the represented gene within the human genome sequence.

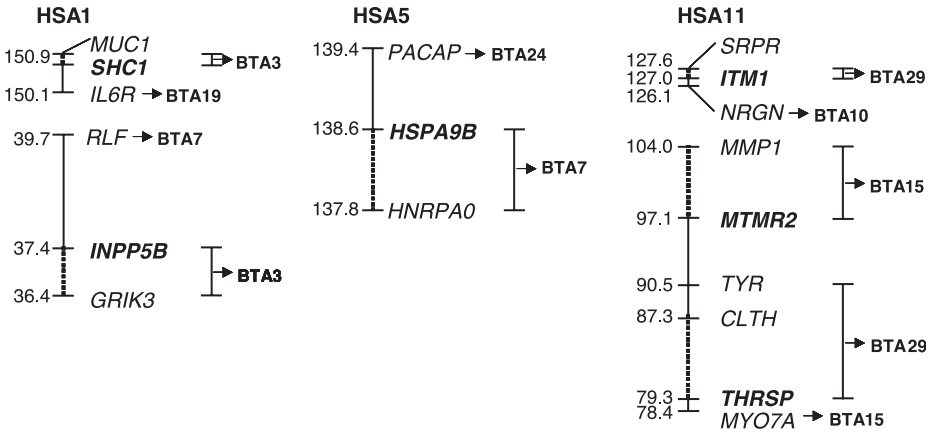

Fig. 1. Extension of conserved synteny between the bovine and the human gene maps by the chromosomal mapping of the *HSPA9B*, *INPP5B*, *ITM1*, *MTMR2*, *SHC1*, and *THRSP* genes in cattle. Distances on human chromosomes are in Mb from p arm telomere of chromosome based on Human Genome Working Draft, November 2002 (http://genome.ucsc.edu/). Dashed lines indicate extended regions of conserved synteny.

In conclusion, this approach was useful in increasing the number of genes positioned on the bovine gene maps and will aid in the integration of the bovine genetic and physical maps. In particular, this study provides map information for genes relevant to mammary gland development and function and may aid in the identification of candidate genes affecting economically important traits in dairy cattle production.

Materials and methods

Primer design

EST from a bovine mammary gland cDNA library [15] were assembled into TC sequences by TIGR (http://www.tigr.org/tdb/btgi/). Those with an expression bias for the mammary gland (>75% of EST in TC) were selected for RH and linkage mapping. Selected TC sequences were aligned with a human genome draft sequence using BLASTN [16] to determine UniGene identity (http://www.ncbi.nlm.nih.gov/UniGene/) and map position. Human genomic sequence structure information was obtained from the human genome browser (http://www.ensembl.org/Homo_sapiens/) or the National Center for Biotechnology Information LocusLink Evidence Viewer (http://www.ncbi.nlm.nih.gov/LocusLink/). Primers were designed for PCR using Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) in regions predicted to amplify a product containing an intron of less than 5 kb. Primer sequences, PCR conditions, and the putative intron of each gene amplified are described in each GenBank dbSTS database entry.

Marker development and linkage mapping

Sequence tagged sites amplified from genomic DNA of the four sires of the USDA bovine reference population [17] were purified using a Montage PCR-96 Cleanup Kit (Millipore, Bedford, MA) and sequenced directly. Sequencing was performed using a CEQ8000 automated sequencer and DTCS Quickstart chemistry (Beckman Coulter, Fullerton, CA). Amplification of the targeted gene region was confirmed using BLASTN, and SNP to be used as markers for linkage mapping were identified using heterozygote detection software (Beckman Coulter CEQ8000 Genetic Analysis System). SNP marker genotypes from progeny of informative sires were characterized primarily by direct sequencing using the CEQ8000 or an ABI Prism 3700 DNA automated sequencer (Applied Biosystems, Foster City, CA). Two gene-associated STS, *CHD4* and *ENTPD6*, were linkage mapped using PCR restriction fragment length polymorphism markers. For *CHD4*, the restriction endonuclease *Rsa*I (New England Biolabs, Beverly, MA) was used for allele determination. Determination of *ENTPD6* alleles was accomplished using the enzyme *Bsg*I (New England Biolabs). Digestion of the 1600-bp PCR product resulted in four alleles designated by restriction fragments of approximately 975 bp (allele 1), 1200 bp (allele 2), 1300 bp (allele 3), and 1100 bp (allele 4). Linkage analysis was performed using CRI-MAP version 2.4 [18,19].

RH mapping

Sizes of PCR amplicons generated from bovine and hamster control DNA were compared by agarose gel electrophoresis for each gene-associated STS. If the bovine fragment was distinguishable from hamster, gene-associated products were amplified by PCR from a panel of 94 clone lines (Bovine/Hamster RH Panel, ResGen Invitrogen, Carlsbad, CA). For *ACTR1A*, *ATIC*, *CAT*, *HSPA9B*, *KRT5*, *LALBA*, *PACSIN2*, *PAPOLG*, and *S100B*, nested primers were designed from STS within regions corresponding to the intron to ensure amplification of a smaller single fragment, clearly distinguishable from hamster. PCR products were separated by agarose gel electrophoresis and concordance data for each cell line were submitted to the Roslin RH database (http://databases.roslin.ac.uk/radhyb/intro.py) for mapping.

Two-point analysis was performed using the algorithm described by Lange and colleagues [20]. Two-point lod scores were evaluated against approximately 1200 microsatellite markers previously mapped to the RH panel [8].

Acknowledgments

The authors acknowledge the technical assistance of Evagelia Laiakis, Amanda Hummel, Laura Ewing, and Marsha Campbell for PCR and sequencing and Jill Philpot for assimilation and submission of sequence data to the GenBank database. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

References

[1] T.S. Sonstegard, et al., Comparative map alignment of BTA27 and HSA4 and 8 to identify conserved segments of genome containing fat deposition QTL, Mamm. Genome 11 (2000) 682–688.
[2] T.E. Scheetz, et al., Generation of a high-density rat EST map, Genome Res. 11 (2001) 497–502.
[3] P. Avner, et al., A radiation hybrid transcript map of the mouse genome, Nat. Genet. 29 (2001) 194–200.
[4] A. Rink, et al., A first-generation EST RH comparative map of the porcine and human genome, Mamm. Genome 13 (2002) 578–587.
[5] Y. Lahbib-Mansais, et al., Contribution to high-resolution mapping in pigs with 101 type I markers and progress in comparative map between humans and pigs, Mamm. Genome 14 (2003) 275–288.
[6] R.T. Stone, et al., Use of bovine EST data and human genomic sequences to map 100 gene-specific bovine markers, Mamm. Genome 13 (2002) 211–215.
[7] M.R. Band, et al., An ordered comparative map of the cattle and human genomes, Genome Res. 10 (2000) 1359–1368.

[8] J.L. Williams, et al., A bovine whole-genome radiation hybrid panel and outline map, Mamm. Genome 13 (2002) 469–474.
[9] M.Y. Hwang, Y.J. Kang, Y.H. Kim, S.W. Scherer, L.C. Tsui, U. Sohn, Generation and chromosome mapping of expressed sequence tags (ESTs) from a human infant thymus, Genome 42 (1999) 457–464.
[10] K. Malone, M.M. Sohocki, L.S. Sullivan, S.P. Daiger, Identifying and mapping novel retinal-expressed ESTs from humans, Mol. Vision 4 (1999) 5–8.
[11] A.K. Fridolfsson, et al., Expansion of the pig comparative map by expressed sequence tags (EST) mapping, Mamm. Genome 8 (1997) 907–912.
[12] C.B. Jorgensen, A.K. Wintero, M. Yerle, M. Fredholm, Mapping of 22 expressed sequence tags isolated from a porcine small intestine cDNA library, Mamm. Genome 8 (1997) 423–427.
[13] R. Davoli, et al., Isolation of porcine expressed sequence tags for the construction of a first genomic transcript map of the skeletal muscle in pig, Anim. Genet. 33 (2002) 3–18.
[14] C. Karall-Albrecht, et al., Mapping of 16 ESTs expressed in the bovine mammary gland during lactation, Mamm. Genome 11 (2000) 320–325.
[15] T.S. Sonstegard, et al., Analysis of bovine mammary gland EST and functional annotation of the *Bos taurus* gene index, Mamm. Genome 13 (2002) 373–379.
[16] S.F. Altschul, et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[17] M.D. Bishop, et al., A genetic linkage map for cattle, Genetics 136 (1994) 619–639.
[18] P. Greene, K. Falls, S. Crooks, CRI-MAP documentation, version 2.4. http://linkage.rockefeller.edu/soft/crimap/, 1990.
[19] E.S. Lander, P. Green, Construction of multilocus genetic linkage maps in humans, Proc. Natl. Acad. Sci. USA 84 (1987) 2363–2367.
[20] K. Lange, M. Boehnke, D.R. Cox, K.L. Lunetta, Statistical methods for polyploid radiation hybrid mapping, Genome Res. 5 (1995) 136–150.